Integration of Transcriptional and Mutational Data improves the Stratification of Peripheral T-Cell Lymphoma

Francesco Maura[1-2]*, Luca Agnelli[2-3]*#, Daniel Leongamornlert[1], Niccolò Bolli[1-2,5], John Chan[4], Anna Dodero[5], Cristiana Carniti[5], Tayla Heavican[6], Alessio Pellegrinelli[7], Giancarlo Pruneri[7], Adam Butler[1], Shriram Bhosle[1], Annalisa Chiappella[8], Alice Di Rocco[9], Pier Luigi Zinzani[10], Francesco Zaja[11], Roberto Piva[12], Giorgio Inghirami[12,13], Wenyi Wang[14], Teresa Palomero[15], Javeed Iqbal[6], Antonino Neri[2-3], Peter J Campbell[1] and Paolo Corradini[2,5#]

[1]Cancer, Ageing and Somatic Mutation Programme, Wellcome Trust Sanger Institute, Hinxton, United Kingdom; [2]Department of Oncology and HematoOncology, University of Milan, Milan, Italy; [3]Hematology, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy;
[4]City of Hope, Duarte, CA;
[5]Department of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy;
[6]Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha;
[7]Department of Pathology and Laboratory Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy;
[8]Department of Hematology, Azienda Ospedaliera Città della Salute e della Scienza, Torino, Italy;
[9]Sapienza University of Rome, Rome, Italy;
[10]Institute of Hematology, University of Bologna, Bologna, Italy
[11]Clinica Ematologica, DAME, University of Udine, Udine, Italy;
[12]Center for Experimental Research and Medical Studies, University of Torino, Department of Molecular Biotechnology and Health Sciences, Torino, Italy;
[13] Pathology and Laboratory Medicines, Weill Cornell Medical College,
[14]The University of Texas MD Anderson Cancer Center, Houston
[15] Institute for Cancer Genetics, Columbia University, New York, NY 10032
*The first two authors equally contributed

Key words: PTCL, Gene expression profile, CIBERSORT, next-generation sequencing

Corresponding Author:
Luca Agnelli
Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy
F. Sforza, 35, 20122 Milan Italy
luca.agnelli@unimi.it
Tel (+39)0255033328
Fax: (+39)0255034571

Abstract

The histological diagnosis of peripheral T-cell lymphoma (PTCL) can represent a challenge, particularly in the case of close related entities like angio-immunoblastic T-lymphomas (AITL) and PTCL-not otherwise specified (PTCL-NOS). Although gene expression profiling and next generations sequencing have been proven to define specific features recurrently associated with distinct entity, genomic-based stratifications have not yet led to definitive diagnostic criteria and/or entered into the routine clinical practice.

Herein, to improve the current molecular classification between AITL and PTCL-NOS, we analyzed the transcriptional profiles from 503 PTCLs stratified according to their molecular configuration and integrated them with genomic data of recurrently mutated genes (RHOA^{G17V} , TET2, IDH2^{R172}, and DNMT3A) in 53 cases (39 AITLs and 14 PTCL-NOSs) included in the series. Our analysis unraveled that the mutational status of *RHOA*, *TET2* and *DNMT3A* poorly correlated, individually, with peculiar transcriptional fingerprint. Conversely, a strong transcriptional signature was identified in *IDH2*^{R172} samples. The integrated analysis of clinical, mutational and molecular data led to 19-genes model whose expression can more precisely differentiate the main PTCL nodal entities. According to the new proposed stratification, *RHOA*^{G17V}, *TET2* and *IDH2*^{R217} mutations emerged as more robust diagnostic markers than previously reported for the differentiation between AITL and PTCL-NOS.

This new gene classifier is simple and reproducible and able to improve the current molecular classification and diagnosis of PTCL.

Introduction

Peripheral T-cell lymphomas (PTCL) represent a heterogeneous group of nodal and extra-nodal mature T-cell neoplasms accounting for approximately 10-15% of all lymphoma in the Western countries.[1-5] The current WHO classification recognizes several distinct subsets, and multiple provisional entities have been recently proposed.[3] The most frequent categories include PTCL-not otherwise specified (PTCL-NOS), angioimmunoblastic lymphoma (AITL), and anaplastic large T cell lymphoma (ALCL) with or without translocations involving *ALK*. [3] Overall, these entities encompass approximately 60% of all PTCL. With the exception of ALCL with *ALK* translocations, PTCL have an aggressive clinical course and poor response to conventional chemotherapy. Although new agents have been recently approved and several others are in clinical trials, the overall clinical success remains unsatisfactory.[1,6-8] This is partially due to a limited knowledge of PTCL biology and reproducible/informative pre-clinical models, which have significantly hindered any mechanistic study aimed at testing candidate targeted treatments.

Even though pathological diagnoses can be rendered in many patients, difficulties can occur, and in particular for those samples sharing features borderline between AITL and PTCL-NOS.[9] Previous studies have shown that these two latter entities might bear distinct gene expression profiles (GEP).[10-16] However, clinical usage of GEP has been limited due to its technical availability and to the absence of a consensus gene signature. Recent advances of next generation sequencing (NGS), and the discovery of recurrent mutated genes (*RHOA*, *TET2*, *DNMT3A*) in approximately 60-70% of AITL and in 20-30% of PTCL-NOS, have somehow changed this landscape.[15,17-20] Notably, 20-30% of AITLs cases can carry hotspot $IDH2^{R172}$ mutations that are virtually absent in PTCL-NOSs.[17] Nevertheless, these findings have not yet significantly impacted the clinical daily practice, which largely relies on histology and phenotype of tumor cells.[3] Moreover, while some mutations appear to be linked to distinct GEP signature(s),[15] the full potential of an integrated genotypic-transcriptomic analysis has not been thoroughly tested.

Here, we collected a large GEP dataset of PTCL, and performed an integrative analysis with mutational data to improve the accuracy of disease classification, aiming of recognizing and more precisely stratifying AITL/PTCL-NOS lymphoma.

Material and Methods

We analysed 503 PTCL and 38 normal T-cell cases, univocally acquired from 8 studies (GSE6338, GSE14879, GSE19067, GSE19069, GSE58445 and GSE65823 at http://www.ncbi.nlm.nih.gov/geo/; ETABM702 and ETABM783 at https://www.ebi.ac.uk/arrayexpress) (**Supplementary Figure 1**).[10,12,21-24] Normalized data were extracted from CEL files using RMA procedure and the annotation for HG-U133Plus2.0 arrays available at http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/21.0.0/entrezg.asp. A batch-effect correction was applied as previously described.[25,26] The whole data set with all available clinical and genomic information acquired was uploaded to https://github.com/emacgene/PTCL.

To investigate the contribution of recurrent mutations to the transcriptional pattern, we adapted a recently published analysis[27] to 39 AITLs and 14 PTCL-NOSs for whom mutational data for *IDH2*^{R172}, *DNMT3A*, *TET2* and *RHOA* were available.[15] Next, we created a model that included the lymphoma histotype, gene expression profile, genotype, and age and gender of each patient. Finally, *ConsensusClusterPlus* package for R [28] was used to determine the significance and robustness of natural grouping of patients based on selected transcriptional data, using Ward and Euclidean as linkage and distance metrics, respectively. CIBERSORT analysis was performed as previously described, using standard procedure and LM22 signature.[29] Contingency analyses were performed by Fisher's exact test (*fisher.test* R function). The CIBERSORT different contribution for each signature was then tested by *pairwise.wilcox.test* R function. Benjamini-Hochberg correction was used for multiple testing adjustment. The pathway enrichment analysis was performed using different modalities. First, the *tmod* R package was used on *limma*-derived signatures to decipher

whether clusters deregulate blood cell-associated transcriptional modules described by Chaussabel *et al.*[30] and by Li *et al.*,[31] according to the procedure described by Weiner et al.[32] The full analysis process written in R is provided in **Supplementary Data 1.** Second, gene set enrichment analysis was run on extended gene set modules using the GSEA java tool from the Broad institute (http://software.broadinstitute.org/gsea/index.jsp). In this second analysis, to reduce the incidence of false positives results, we choose the *samr* package, and a more conservative approach, to list differentially expressed genes.

To validate our findings in an independent dataset, we imported data from 34 previously published RNAseq samples (dbGap accession n. phs000689.v1.p1).[17] The RNAseq raw expression data were normalized as previously described.[33]The full computational process written in R is provided in **Supplementary Data 2**.

Results

Dataset definition

We analysed the transcriptional profiles of 503 PTCL cases and 38 normal donors provided by 8 different studies.[10,12,13,21-24,34] According to the most recent updates in the T-cell lymphoma molecular classification,[3] these series included the following histotypes: 127 AITL, 144 PTCL-NOS, 56 *ALK*+ ALCL, 96 *ALK*- ALCL, 21 Adult T-Cell Lymphoma (ATLL), 59 NK/T-cell lymphomas (**Figure 1a**). By unsupervised hierarchical clustering and principal component analysis, *ALK*- and *ALK*+ ALCLs, ENKTL and ATLL were associated with distinct signatures; conversely, the transcriptional portrait of AITL and PTCL-NOS displayed a considerable overlap (**Figure 1b-c**).

IDH2^{R172} but not *RHOA*, *TET2* and *DNMT3A* is associated with specific transcriptional pattern

To improve the distinction between AITL and PTCL-NOS, we integrated expression profiles with clinical, laboratory and mutational data (39 AITLs and 14 PTCL-NOSs) (See **Material and Methods** and **Supplementary Data 1**). This strategy lead to a preliminary list of 221/20.363 (1%) differentially modulated genes [false rate discovery (FDR) <1%] **(Figure 2a** and **Supplementary Table 1**). Among these, 30 of them emerged as significantly and specifically associated to one distinct variable. Gender selectively impacted 14 genes located on the X and Y chromosome and for this reason they were not included for further analysis. Interestingly, the mutational status of *RHOA*^{G17V}, *TET2* and *DNMT3A* poorly correlated with any distinct transcriptional profile in both AITL and PTCL-NOS, as no gene showed differential expression to be correlated with the genotype (**Figure 2b**). Conversely, *IDH2*^{R172} samples were linked to a unique transcriptional signature including 3 genes with FDR <1%: *ID2*, *NETO2* and *SLC5A3* (**Supplementary Data 1** and **Figure 2b**). Notably, AITL and PTCL-NOS transcriptional profiles could be differentiated by the expression of 13 genes, including 3 of them recently described in recent gene expression model (*ROBO1*, *ARHGEF10* and *EFNB2*) (**Figure 2a-b**).[12]

Lastly, we decided to combine the signature specifically associated with *IDH2*^{R172} mutational status (n=3) and those linked to histological diagnoses (n=13) to all AITL and PTCL-NOS (n=271) samples. This approach led to a hierarchical clustering with an improved clustering of PTCL-NOS and AITL cases. Collectively, these data are consistent with biologically distinct subgroups associated with specific mutational-based signatures, and linked to defined pathological features (**Figure 2c**).

A 19-gene expression-based model distinguished PTCL subtypes

Our group has previously described a 3-gene signature capable to effectively distinguish *ALK*- ALCLs from PTCL-NOS[35]. Here, we combined this 3-gene signature with our new 16-gene model, ultimately leading to a 19-gene model, aimed at improving the stratification of all major PTCL entities (*ALK*- ALCL, PTCL-NOS and AITL). This led to the recognition of five distinct subsets (**Figure 3, Supplementary Data 1** and **Table 1**). The first group (C-1; n=87) was mainly composed by AITL samples (93%; 81/87), and enriched by *IDH2*^{R172} (16/36; 44%) and *RHOA*^{G17V}(23/29; 79%) mutated lymphoma. The second (C-2; n=103), with an AITL-like signature, included samples annotated as well as PTCL-NOSs (64/103; 62%) and AITLs (33/103; 32%). In this group, *IDH2*^{R172} samples were minimally represented compared to C-1 (1/23; p=0.001); conversely, a higher prevalence of *RHOA* ^{G17V} mutations was observed (15/29; 52%); particularly higher than that seen in cluster 4 (p=0.005). The third cluster represented only by 21 patients, included samples that could be linked to any of 3 main PTCL entities. Group 4 (C-4; n=63) included largely PTCL-NOSs (52/63; 82%), displaying a distinctive transcriptional profile characterized by a low prevalence of *RHOA* and *TET2* mutations compared to C-1 (p<0.0001 and p=0.0002 respectively) or C-2 (p=0.005 and 0.03 respectively) (**Supplementary Figure 2**).[12] In the last cluster (C-5; n=55) *ALK*- ALCL cases (47/69; 68%) were over-represented, confirming the strong association between these lymphoma and the expression of *TNFRSF8*, *BATF3* and *TMOD1*.[35] No significant differences were observed in the *DNMT3A* mutational profile between any cluster (**Table 1**). Overall, this approach could improve the previous molecular

characterization of up to 8% AITLs, 31% *ALK*- ALCL and 29% PTCL-NOSs providing the bases of the recognition of unique biological subgroups (**Table 1**).

Next, to discover potential gene classifiers, we investigated the global gene expression among samples within the 5 clusters (**Supplementary Data 1**). Firstly, we confirmed molecular signatures previously associated to AITL and PTCL-NOS (i.e *ROBO1, LPAR1, SOX8, TUBB2B, TNFRSF8, TMOD1, BATF3, ATP6V0D1, CHI3L1, CREG1, CTSB, CTCS, FTL, HCK*), strengthening the ability of the 19-gene model to differentiate distinct pathological entities.[12,36]

Running 2 independent pathway analyses (**Material and Methods and Supplementary Data 1**), we observed a significant enrichment of B-cell and Plasma cell pathways among C-1 and C-2, in agreement with the notion that AITL are often enriched in B-cells/plasma cells; conversely the cluster 5 showed an enrichment for T-CD4 cell cycles pathway, but a significant low involvement of ITK/PKC and T-cell activation pathway (**Supplementary Figure 3**), in agreement with the consideration that the TCR signalling has been reported to be non-dominant in this group.[37]

Validation Cohort

To validate our 19-gene signature, we analyzed the RNAseq expression of 34 PTCL-NOSs (11 AITL, 11 PTCL-NOS, 8 *ALK*- ALCL and 4 *ALK*+ ALCL) from a previously published dataset.[17] The expression data of 2 of 19 genes were not available (AL441992.1 and SLC5A3). Using the remaining 17 genes we were able to extract 4 main clusters (**Supplementary Data 2** and **Figure 4**). The first cluster was compatible with the C-1 and C-2 (AITL and AITL-like) clusters. All *RHOA*, *TET2* and $IDH2^{R172}$ mutated cases were included in this cluster mostly composed by a fraction of PTCL-NOSs (6/11) and all AITLs (10/10). The second was characterized by a profile similar to the C-4 group, with PTCL-NOS without *RHOA* and *TET2* mutations. The remaining clusters included the great majority of *ALK*+ and *ALK*- ALCL cases (9/11).

Computational investigation defines distinct tumor-host compositions

Seeking to discover whether the new PTCL clusters may be linked/associated to any host-tumor features, we used CIBERSORT to define the cellular composition of the tumor and host elements. CIBERSORT is a computational tool that accurately resolves relative fractions of different elements within complex environments (**Figure 5a** and **Supplementary Data 1**).[29] This analysis revealed that the C-1 and C-2 clusters were characterized by a highest contribution of plasma cell and B-cell signatures (**Figure 5b**). Conversely, both clusters displayed a very low T-reg but high macrophages M_0 signals. Cluster 5, mainly represented by ALK- ALCLs, was enriched for activated T-cell CD4+ and by Macrophages (M_1 and M_2) signatures (**Figure 5b** and **Supplementary Data 1**). Notably, the contribution of T-follicular helper cells was significantly lower in C5 than in the C-1, C-2 and C-4 groups. Lastly, the C-3 cluster was characterized by a unique profile associated with NK-cell ($p<0.01$) and activated mast cell ($p<0.001$) signatures, likely reflecting unique features of these lymphoma compared to those seen on bona fide AITL, PTCL-NOS and ALCL cases (**Supplementary Data 1**).

Discussion

GEP has emerged as one of the most robust and reliable approach to differentiate human lymphoma. This has also been true for T-cell lymphoma, a strategy that have allowed the identification of even closely related entities like AITL and PTCL-NOS. [10,12,22-24,34,35] However, GEP profile from formalin fixed paraffin embedded samples has been technical problematic and thus the molecular/expression stratification of PTCL has not entered yet into the routine clinical practice. Recently, novel technologies (e.g. NanoString) have provided for the first time a reproducible and feasible quantification of specific transcripts, mainly for diffuse large B-cell lymphoma.[38-40] conversely, this approach has not yet been tested in PTCL, likely because a short and robust list of differentially expressed genes has not emerged yet.

To bridge this gap, we performed meta-analysis of the largest PTCL series collected to date to assess the molecular profile of the main PTCL subgroups and define a usable list of significant differentially expressed genes. By integrating NGS and GEP data, we were able to discover the transcriptional impact of the recurrent mutations of AITLs and PTCL-NOSs. Indeed, we found that *TET2*, *DNMT3A*, *RHOA*^{G17V} mutations did not show any distinct gene expression signatures. This last mutation was detected at sub-clonal level in a significant fraction of these cases and this may explain its low impact on transcriptional configuration. [15] *TET2* and *DNMT3A* mutations are likely as early as hematopoietic stem cells mutations involved in clonal hemopoiesis;,[41,42] therefore, we could not exclude that *TET2*-associated gene expression signature might be shared by other subtypes and masked by AITL molecular classification, subsequently limiting its extraction through our statistical process. Conversely, in line with a previous report,[15] the *IDH2*^{R172} mutation significantly correlated with a distinct expression signature independent from the molecular subgroup. By integrating pathology, NGS and GEP analyses we defined a short list of highly predictive genes (n=19) whose differential expression divided the samples into 5 clusters, strongly associated with distinct PTCL entities. Specifically, C-1 and C-2 were characterized by most of the AITL hallmarks. In addition, despite their limited gene expression impact, the great

majority of *RHOA* ^{G17V} and *TET2* mutations were grouped together, confirming their potential utility in the diagnostic process of these lymphoma. A significant fraction of PTCL-NOS were included in these 2 AITL clusters sharing high prevalence of *RHOA* and *TET2* mutations and expressing some distinct, and in part previously reported, AITL-associated genes, suggesting a better ability of this model to re-assign more correctly the T-follicular helper PTCLs.

Furthermore, even without a direct GEP impact, *RHOA* mutations were mostly clustered within AITL clusters, emerging together with *IDH2*^{R217} mutations as a potential specific diagnostic marker to be further employed in routine diagnostics. Lastly, The majority of PTCL-NOS not assigned to C-2, composed a unique and distinct cluster (C4).

Based both on the present and recently published[10,12,37] data, PTCL-NOS group within the C-4 could not be considered any more like an “orphan” or “not-otherwise specified” subset, but instead as a distinct biologically entity. In fact, the lymphoma patients within this cluster have a distinct genotypic and transcriptomic pattern, mostly characterized by a significant enrichment in downregulated genes, confirming previous findings.[13] Clearly, further studies will be needed to confirm our findings, and to further refine the list of genomic driver events and genes whose expression levels may have diagnostic value. However, our data were preliminarily cross-validated in an independent small RNAseq cohort[17], confirming the robustness of the signature. This was true even across gene expression platforms, and thus represents a promising test bed for future routine nanostring approaches.

Combining the expression levels and the mutational status of a limited set of genes in future studies, we will be able to improve the current diagnostic approach of PTCLs and lay the basis for effective treatments through the identification of recurrently dysregulated pathways.

Acknowledgment

NB is funded by AIRC (Associazione Italiana per la Ricerca sul Cancro) through a MFAG (n.17658). FM is supported by A.I.L. (Associazione Italiana Contro le Leucemie-Linfomi e Mieloma ONLUS)

Authorship Contributions:

F.M., L.A. and P.C.: designed the study, collected and analysed the data, and wrote the paper;; D.L., W.W., P.J.C., and S.B. analysed the data.; J.C., T.H., A.D., C.C., A.P., G.P., A.B., A.C., A.D.R., P.L.Z., F.Z., U.V., R.P., J.I., T.P. and J.I. collected the data; N.B., A.N., and J.I. critically revised the paper

Disclosure of Conflicts of Interest:

No conflict of interests to declare

Figure Legends:

Figure 1. a) Molecular composition of the gene expression cohort (541 cases). Unsupervised hierarchical clustering (b) and PCA analysis (c) on the entire series.

Figure 2. a) Distribution of the variance of expression levels across genes explained by clinical, molecular and genetic alterations (f-test; FDR<1%; n=221). b) Statistically significant mutation expression interaction terms (f-test; FDR<1%), for each alteration and clinical variable. The associated logarithmic expression fold change is indicated by colour. c) Hierarchical clustering of 271 PTCL-NOSs and AITLs based on the 16-signfiicantly extracted genes. The mutational status for *TET2*, *RHOA*, *DNMT3A, IDH2*^{R172} was reported on the top.

Figure 3. Heatmap of the 19-genes model including all PTCL-NOS, AITL and *ALK*- ALCL cases (n=367), stratified according to the cluster determined by the *ConsensusClusterPlus* R function.

Figure 4. a) CIBERSORT profile of each enrolled tumour sample (n=503). b) Boxplot of the relative estimated percentage for the most relevant and significant cell types, stratified by groups (Wilcoxon rank sum test for comparisons).

Figure 5. Validation of the 19-gene model on a RNAseq external series.

Supplementary Figure Legend

Supplementary Figure 1. Summary of the histological composition of each external gene expression data set included in this study.

Supplementary Figure 2. Heat map based on the differential expression of the 16-genes between the C-2 and C-4 PTCL-NOSs.

Supplementary Figure 3. The *tmod* most significant differentially enriched pathway among the extracted 5 clusters,

References

1 Ellin, F, Landstrom J, Jerkeman M & Relander T. Real-world data on prognostic factors and treatment in peripheral T-cell lymphomas: a study from the Swedish Lymphoma Registry. *Blood* 2014; **124**:1570-1577.

2 Maura F, DA, Carniti C, Bolli N. Biology of Peripheral T cell lymphomas - not otherwise specified: is something finally happening? *Pathogenesis* 2016; **3** 9–18.

3 Swerdlow, SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R *et al.* The 2016 revision of the World Health Organization (WHO) classification of lymphoid neoplasms. *Blood* 2016.

4 Van Arnam, JS, Lim MS & Elenitoba-Johnson KSJ. Novel insights into the pathogenesis of T-cell lymphomas. *Blood* 2018; **131**:2320-2330.

5 Vose, J, Armitage J, Weisenburger D & International TCLP. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. *J Clin Oncol* 2008; **26**:4124-4130.

6 Corradini, P, Vitolo U, Rambaldi A, Miceli R, Patriarca F, Gallamini A *et al.* Intensified chemo-immunotherapy with or without stem cell transplantation in newly diagnosed patients with peripheral T-cell lymphoma. *Leukemia* 2014; **28**:1885-1891.

7 d'Amore, F, Relander T, Lauritzsen GF, Jantunen E, Hagberg H, Anderson H *et al.* Up-front autologous stem-cell transplantation in peripheral T-cell lymphoma: NLG-T-01. *J Clin Oncol* 2012; **30**:3093-3099.

8 Gallamini, A, Stelitano C, Calvi R, Bellei M, Mattei D, Vitolo U *et al.* Peripheral T-cell lymphoma unspecified (PTCL-U): a new prognostic model from a retrospective multicentric clinical study. *Blood* 2004; **103**:2474-2479.

9 Herrera, AF, Crosby-Thompson A, Friedberg JW, Abel GA, Czuczman MS, Gordon LI *et al.* Comparison of referring and final pathology for patients with T-cell lymphoma in the National Comprehensive Cancer Network. *Cancer* 2014; **120**:1993-1999.

10 Iqbal, J, Weisenburger DD, Greiner TC, Vose JM, McKeithan T, Kucuk C *et al.* Molecular signatures to improve diagnosis in peripheral T-cell lymphoma and prognostication in angioimmunoblastic T-cell lymphoma. *Blood* 2010; **115**:1026-1036.

11 Iqbal, J, Wilcox R, Naushad H, Rohr J, Heavican TB, Wang C *et al.* Genomic signatures in T-cell lymphoma: How can these improve precision in diagnosis and inform prognosis? *Blood Rev* 2015.

12 Iqbal, J, Wright G, Wang C, Rosenwald A, Gascoyne RD, Weisenburger DD *et al.* Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. *Blood* 2014; **123**:2915-2923.

13 Piccaluga, PP, Fuligni F, De Leo A, Bertuzzi C, Rossi M, Bacci F *et al.* Molecular profiling improves classification and prognostication of nodal peripheral T-cell lymphomas: results of a phase III diagnostic accuracy study. *J Clin Oncol* 2013; **31**:3019-3025.

14 Piva, R, Agnelli L, Pellegrino E, Todoerti K, Grosso V, Tamagno I *et al.* Gene expression profiling uncovers molecular classifiers for the recognition of anaplastic large-cell lymphoma within peripheral T-cell neoplasms. *J Clin Oncol* 2010; **28**:1583-1590.

15 Wang, C, McKeithan TW, Gong Q, Zhang W, Bouska A, Rosenwald A *et al.* IDH2R172 mutations define a unique subgroup of patients with angioimmunoblastic T-cell lymphoma. *Blood* 2015; **126**:1741-1752.

16 Wang, T, Feldman AL, Wada DA, Lu Y, Polk A, Briski R *et al.* GATA-3 expression identifies a high-risk subset of PTCL, NOS with distinct molecular and clinical features. *Blood* 2014; **123**:3007-3015.

17 Palomero, T, Couronne L, Khiabanian H, Kim MY, Ambesi-Impiombato A, Perez-Garcia A *et al.* Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nat Genet* 2014; **46**:166-170.

18 Sakata-Yanagimoto, M, Enami T, Yoshida K, Shiraishi Y, Ishii R, Miyake Y *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nat Genet* 2014; **46**:171-175.

19 Schatz, JH, Horwitz SM, Teruya-Feldstein J, Lunning MA, Viale A, Huberman K *et al.* Targeted mutational profiling of peripheral T-cell lymphoma not otherwise specified highlights new mechanisms in a heterogeneous pathogenesis. *Leukemia* 2015; **29**:237-241.

20 Yoo, HY, Sung MK, Lee SH, Kim S, Lee H, Park S *et al.* A recurrent inactivating mutation in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nat Genet* 2014; **46**:371-375.

21 Eckerle, S, Brune V, Doring C, Tiacci E, Bohle V, Sundstrom C *et al.* Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to Hodgkin lymphoma. *Leukemia* 2009; **23**:2129-2138.

22 Iqbal, J, Weisenburger DD, Chowdhury A, Tsai MY, Srivastava G, Greiner TC *et al.* Natural killer cell lymphoma shares strikingly similar molecular features with a group of non-hepatosplenic gammadelta T-cell lymphoma and is highly sensitive to a novel aurora kinase A inhibitor in vitro. *Leukemia* 2011; **25**:348-358.

23 Piccaluga, PP, Agostinelli C, Califano A, Carbone A, Fantoni L, Ferrari S *et al.* Gene expression analysis of angioimmunoblastic lymphoma indicates derivation from T follicular helper cells and vascular endothelial growth factor deregulation. *Cancer Res* 2007; **67**:10703-10710.

24 Scarfo, I, Pellegrino E, Mereu E, Kwee I, Agnelli L, Bergaggio E *et al.* Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. *Blood* 2016; **127**:221-232.

25 Johnson, WE, Li C & Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007; **8**:118-127.

26 Leek, JT, Johnson WE, Parker HS, Jaffe AE & Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; **28**:882-883.

27 Gerstung, M, Pellagatti A, Malcovati L, Giagounidis A, Porta MG, Jadersten M *et al.* Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat Commun* 2015; **6**:5901.

28 Wilkerson, MD & Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010; **26**:1572-1573.

29 Gentles, AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med* 2015; **21**:938-945.

30 Chaussabel, D, Quinn C, Shen J, Patel P, Glaser C, Baldwin N *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* 2008; **29**:150-164.

31 Li, S, Rouphael N, Duraisingham S, Romero-Steiner S, Presnell S, Davis C *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* 2014; **15**:195-204.

32 Weiner, J, 3rd, Parida SK, Maertzdorf J, Black GF, Repsilber D, Telaar A *et al.* Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS One* 2012; **7**:e40221.

33 Love, MI, Anders S, Kim V & Huber W. RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Res* 2015; **4**:1070.

34 Piccaluga, PP, Agostinelli C, Califano A, Rossi M, Basso K, Zupo S *et al.* Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J Clin Invest* 2007; **117**:823-834.

35 Agnelli, L, Mereu E, Pellegrino E, Limongi T, Kwee I, Bergaggio E *et al.* Identification of a 3-gene model as a powerful diagnostic tool for the recognition of ALK-negative anaplastic large-cell lymphoma. *Blood* 2012; **120**:1274-1281.

36 Wang, M, Zhang S, Chuang SS, Ashton-Key M, Ochoa E, Bolli N *et al.* Angioimmunoblastic T cell lymphoma: novel molecular insights by mutation profiling. *Oncotarget* 2017; **8**:17763-17770.

37 Inghirami, G, Chan WC, Pileri S & malignancies AxcG-dtmol. Peripheral T-cell and NK cell lymphoproliferative disorders: cell of origin, clinical and pathological implications. *Immunol Rev* 2015; **263**:124-159.

38 Reddy, A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* 2017; **171**:481-494 e415.

39 Scott, DW, Mottok A, Ennishi D, Wright GW, Farinha P, Ben-Neriah S *et al.* Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. *J Clin Oncol* 2015; **33**:2848-2856.

40 Veldman-Jones, MH, Lai Z, Wappett M, Harbron CG, Barrett JC, Harrington EA *et al.* Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System. *Clin Cancer Res* 2015; **21**:2367-2378.

41 Genovese, G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 2014; **371**:2477-2487.

42 Jaiswal, S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 2014; **371**:2488-2498.